ADDITIVE WATERMARK DETECTORS BASED ON A NEW HIERARCHICAL SPATIALLY ADAPTIVE IMAGE MODEL

Antonis Mairgiotis[1], Nikolaos Galatsanos[1], and Yongi Yang[2]

[1]Dept. of Computer Science
University of Ioannina
Ioannina, Greece 45110
{*mairgiot, galatsanos@cs.uoi.gr*}

[2]Dept. of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, IL 60616, USA
{*yy@ece.iit.edu*}

ABSTRACT

In this paper we propose a new family of watermark detectors for additive watermarks in digital images. These detectors are based on a recently proposed two-level, hierarchical image model, which was found to be beneficial for image recovery problems. The top level of this model is defined to exploit the spatially-varying local statistics of the image, while the bottom level is used to characterize the image variations along two principal directions. Based on this model we derive a class of detectors for the additive watermark detection problem, including the generalized likelihood ratio test (GLRT) and Rao detectors.

Index Terms— watermarking, statistical methods, source modeling

1. INTRODUCTION

Additive watermark detection can be formulated as a hypothesis testing problem, where one needs to determine the presence or absence of a known watermark in an image. Within such a formulation, the watermark is treated as the known signal and the image is treated as the corrupting noise [1]. To derive a test statistic for this problem, such as the likelihood ratio test detector, a statistical model for the image has to be defined.

In this paper we propose the use of a hierarchical, locally adaptive image model for watermark detection. The top level of this model is defined to exploit the spatially-varying local statistics of the image. This model can be viewed as a generalization of the concept of line process used in the context of compound Markov random fields [6]. The difference is that a continuous model, rather than binary edges, is used for characterizing the local discontinuities in the image. Using this image model we will derive detectors for additive watermarking, which include the generalized likelihood ratio test (GLRT) and Rao detectors.

We note that the hierarchical image model used in this study was recently developed for image restoration in [5]. It is interesting to note that the development of image models has also been very important for the classical image denoising and restoration problems, in which a statistical image model is essential for various estimation methodologies, e.g., maximum a posteriori estimation.

The rest of this paper is organized as follows. In Section 2 we introduce the hierarchical image model and formulate its use for additive watermark detection. In Section 3 the GLRT detector is derived and methods to estimate the necessary parameters of the model are given. Rao test based detector is derived in Section 4. In Section 5 numerical experiments are given to demonstrate the proposed detectors.

2. IMAGE MODELS AND ADDITIVE WATERMARKS

The image model we propose to use in this paper is based on the first order differences of the image along the two principal directions. Specifically, consider an image $\mathbf{f}$, whose pixels are denoted by $f(i,j)$. At pixel location (i,j), we define the image directional differences (IDD) along the horizontal and vertical directions, respectively, as follows:

$$\begin{aligned} \varepsilon_1(i,j) &= f(i,j) - f(i,j+1), \\ \varepsilon_2(i,j) &= f(i,j) - f(i+1,j) \end{aligned} \tag{1}$$

We assume that these IDDs obey a Gaussian probability density function (pdf), given by

$$\varepsilon_k(i,j) \sim N\left(0, a_k^{-1}(i,j)\right), \tag{2}$$

where $a_k^{-1}(i,j)$ is the variance parameter. Since for each spatial location different variances are assumed this model is very flexible to model the fluctuations of the IDDs in both smooth and edge image areas.

For notational convenience, in the rest of the paper we will denote the IDDs using a single index as $\boldsymbol{\varepsilon}_k = \left[\varepsilon_k(1), \varepsilon_k(2), ..., \varepsilon_k(N)\right]^T$, k=1,2, where N is the total

number of image pixels. In addition, let $\tilde{\boldsymbol{\varepsilon}} = \left[\boldsymbol{\varepsilon}_1^T, \boldsymbol{\varepsilon}_2^T\right]^T$, a vector consisting of IDDs in both directions.

Assuming independence of the IDDs, we can write the joint pdf

$$p(\tilde{\boldsymbol{\varepsilon}};\tilde{\mathbf{a}}) \propto \prod_{k=1}^{2}\prod_{i=1}^{N}\left[a_k^{1/2}(i)\exp\left(-\frac{1}{2}a_k(i)\left(\varepsilon_k(i)\right)^2\right)\right] \quad (3)$$

where $\tilde{\mathbf{a}} = \left[\mathbf{a}_1^T, \mathbf{a}_2^T\right]^T$, $\mathbf{a}_k = \left[a_k(1), a_k(2), ..., a_k(N)\right]^T$, which denotes the corresponding variance parameters.

The pdf in (3) allows the local variance to vary from pixel to pixel. This is desirable for modeling the spatially non-stationary properties of the image (e.g., edges). Unfortunately, it includes as many variance parameters $a_k(i)$ as the number of image pixels. To avoid the problem of over-fitting, it is necessary to impose additional constraint on the model to limit its degrees of freedom. For this purpose we model $a_k(i)$ as random variables, and define a hyper-prior on them.

In this work we use a Gamma pdf for the hyper-prior, which is of the form

$$p(a_k(i);m,l) \propto a_k^{\frac{l-2}{2}}(i)\exp\left\{-m(l-2)a_k(i)\right\},\ k=1,2, \quad (4)$$

where m and l are the parameters of the Gamma. Such a choice is motivated by the fact that the Gaussian and the Gamma families are conjugate [7] with respect to the inverse of the variance of the Gaussian, of which the benefit will become clear later in the estimate of the model parameters. This combination has also been used successfully in sparse Bayesian models for machine learning tasks.

For the Gamma pdf in (4), we have

$$E\left[a_k(i)\right] = l\left(2m(l-2)\right)^{-1},\ Var\left[a_k(i)\right] = l\left(2m^2(l-2)^2\right)^{-1}$$

Assuming that $a_k(i)$ are independent and identically distributed, then we have

$$p(\tilde{\mathbf{a}};m,l) = C\cdot\prod_{k=1}^{2}\prod_{i=1}^{N}\left(a_k^{\frac{l-2}{2}}(i)\ \exp\left\{-m(l-2)a_k(i)\right\}\right) \quad (5)$$

where C is a normalization constant.

In the additive watermark detection problem one has to decide between the following two hypotheses

$$\begin{aligned} H_0 &: \mathbf{y} = \mathbf{f} \\ H_1 &: \mathbf{y} = \mathbf{f} + \gamma\,\mathbf{w} \end{aligned} \quad (6)$$

where $\mathbf{y}$ and $\mathbf{f}$ are the observed and the original images, respectively, and $\mathbf{w}$ is the watermark signal and γ is its strength.

Applying the directional difference operators $\mathbf{Q}_k$, $k=1,2$, to the observed image in (6), we obtain

$$\begin{aligned} H_0 &: \mathbf{y}_k^{'} = \boldsymbol{\varepsilon}_k \\ H_1 &: \mathbf{y}_k^{'} = \boldsymbol{\varepsilon}_k + \gamma\mathbf{w}_k' = \boldsymbol{\varepsilon}_k + \mathbf{w}_k'' \end{aligned} \quad (7)$$

where $\mathbf{y}_k^{'} = \mathbf{Q}_k\mathbf{y}$, $\boldsymbol{\varepsilon}_k = \mathbf{Q}_k\mathbf{f}$, $\mathbf{w}_k'' = \gamma\mathbf{Q}_k\mathbf{w}$, $k=1,2$.

Based on Eqs. (3) and (5), the conditional pdfs of the observations for the two hypotheses in Eq. (7) can be written as

$$\begin{aligned} &p(\tilde{\mathbf{y}}';\tilde{\mathbf{a}},H_0) = \\ &C\cdot\left\{\prod_{k=1}^{2}\prod_{i=1}^{N}a_k(i)^{1/2}\right\}\exp\left(-\frac{1}{2}\sum_{k=1}^{2}\sum_{i=1}^{N}a_k(i)\left(y_k'(i)\right)^2\right) \\ &p(\tilde{\mathbf{y}}';\tilde{\mathbf{a}},H_1) = \\ &C\cdot\left\{\prod_{k=1}^{2}\prod_{i=1}^{N}a_k(i)^{1/2}\right\}\exp\left(\frac{1}{2}\sum_{k=1}^{2}\sum_{i=1}^{N}a_k(i)\left(y_k'(i)-w_k''(i)\right)^2\right) \end{aligned} \quad (8)$$

where $\tilde{\mathbf{y}}' = \left[\mathbf{y}_1'^T, \mathbf{y}_2'^T\right]^T$. In what follows these pdfs will be used to derive the GLRT and Rao test detectors.

3. GENERALIZED LIKELIHOOD RATIO DETECTOR

The GLRT is given by

$$GLRT(\tilde{\mathbf{y}}') = \log\left\{\frac{p(\tilde{\mathbf{y}}';\hat{\tilde{\mathbf{a}}}_{/H_1},H_1)}{p(\tilde{\mathbf{y}}';\hat{\tilde{\mathbf{a}}}_{/H_0},H_0)}\right\} \underset{H_0}{\overset{H_1}{\gtrless}} 0, \quad (9)$$

where $\hat{\tilde{\mathbf{a}}}_{/H_0}$, $\hat{\tilde{\mathbf{a}}}_{/H_1}$ are the estimates of $\tilde{\mathbf{a}}$ in Eq. (9) under the two hypotheses.

With the conditional pdfs in Eq. (8), the test statistic for the detector in Eq. (9) can be written as

$$\begin{aligned} T(\tilde{\mathbf{y}}') &= \sum_{k=1}^{2}\sum_{i=1}^{N}y_k'(i)^2\left(\hat{a}_k(i)_{/H_0} - \hat{a}_k(i)_{/H_1}\right) \\ &-\sum_{k=1}^{2}\sum_{i=1}^{N}\hat{a}_k(i)_{/H_1}w_k''(i)^2 - 2y_k'(i)w_k''(i)\hat{a}_k(i)_{/H_1} \end{aligned} \quad (10)$$

For weak watermarks, it is reasonable to expect that the estimates $\hat{a}_k(i)_{/H_0}, \hat{a}_k(i)_{/H_1}$ are approximately equal. Thus, the test statistic in Eq. (10) can be simplified as (upon ignoring the middle term as it does not depend directly on the data)

$$T_{GLRT}(\tilde{\mathbf{y}}') = \sum_{k=1}^{2}\sum_{i=1}^{N}y_k'(i)w_k''(i)\hat{a}_k(i)_{/H_1} \underset{H_0}{\overset{H_1}{\gtrless}} T, \quad (11)$$

where T is a threshold that determines the false alarm vs. probability of detection tradeoff of the detector [8].

The simplified test statistic in (11) offers a rather informative insight on the GLRT detector. It assumes essentially the form of a matched filter, where the observation at each pixel is normalized by its local variance.

The test statistic in (10) and (11) requires the estimates of the parameters $\hat{a}_k(i)$. Obviously, the ML estimate here will be problematic because only one data point is available. Instead, we use a Bayesian estimate instead, where the hyper-prior $p(\tilde{\mathbf{a}};m,l)$ is used to ameliorate this difficulty. By invoking the Bayes' law, this estimate is obtained as

$$\hat{a}_k(i)_{/H_0} = \arg\max_{a_k(i)} \left\{ \log p\left(\tilde{\mathbf{a}} \mid \tilde{\mathbf{y}}', H_0, m, l\right) \right\} =$$
$$\arg\max_{a_k(i)} \left\{ \log p\left(\tilde{\mathbf{y}}' \mid \tilde{\mathbf{a}}, H_0\right) + \log p\left(\tilde{\mathbf{a}}; m, l\right) \right\},$$

and similarly for $\hat{a}_k(i)_{/H_1}$. After some algebra, it can be shown that

$$\hat{a}_k(i)_{/H_0} = \frac{1+(l-2)}{(y_k'(i))^2 + 2m(l-2)},$$
$$\text{and} \quad \hat{a}_k(i)_{/H_1} = \frac{1+(l-2)}{(y_k'(i) - w_k''(i))^2 + 2m(l-2)} \quad . \tag{12}$$

It is interesting to examine the effect of the parameter l in this estimate. As $l \to \infty$ the estimate becomes $\hat{a}_k(i) = (2m)^{-1}$ for both H_0 and H_1. That is, the prior dominates the estimate. On the other hand, as $l \to 2$, the prior parameters disappear in (12), and the estimate simply degenerates to the ML estimate. For $l \in (2, \infty)$, the prior "regularizes" the estimate $\hat{a}_k(i)$ where the ML estimate is unstable because of lack of data. In our experiments we tested both the GLRT detector in Eq. (10) and its approximate in (11), and found that their performance was almost identical. For this purpose in the rest of this paper we will report results with the simplified one in Eq. (11).

4. RAO DETECTOR

Thus far in deriving the watermark detectors we have considered the situations that the watermark strength is exactly known (i.e., parameter γ in Eq. (6)). There are also situations it might not be known, e.g., in public watermarking. In such a case one could treat γ the same way as other model parameters and use its estimate in the GLRT detector. However, this becomes problematic in applications where the watermark signal is much weaker than the cover image. Our experiments indicate that this can greatly compromise the accuracy of the ML estimate of γ. In order to address this difficulty, we use the Rao test, which is a locally optimal detector (LOD) with performance close to that of a clairvoyant GLRT (when γ is small) [9]. This detector was first introduced to the image watermarking problem in [4].

Supposing $p'\left(y_k'(i); a_k(i), H_0\right)$ is the derivative of the pdf with respect to the observations [9], the Rao test for the observations in Eq. (7) is given by

$$T_R\left(\tilde{\mathbf{y}}'; \tilde{\mathbf{a}}\right) =$$
$$\frac{2N\left[\sum_{k=1}^{2}\sum_{i=1}^{N} w_k'(i) \frac{p'\left(y_k'(i); a_k(i), H_0\right)}{p\left(y_k'(i); a_k(i), H_0\right)}\right]^2 \underset{H_0}{\overset{H_1}{\gtrless}} T}{\left[\sum_{k=1}^{2}\sum_{i=1}^{N} w_k'(i)\right]^2 \left[\sum_{k=1}^{2}\sum_{i=1}^{N} \frac{p'\left(y_k'(i); a_k(i), H_0\right)}{p\left(y_k'(i); a_k(i), H_0\right)}\right]^2} \tag{13}$$

It is noted that in Eq. (14) it is only the watermark shape $\mathbf{w}_k'$ (not the parameter γ) that is necessary for the Rao detector. Substituting the pdfs model into previous equation, we obtain

$$T_R\left(\tilde{\mathbf{y}}'; \hat{\tilde{\mathbf{a}}}_{/H_0}\right) =$$
$$\frac{\left[\sum_{k=1}^{2}\sum_{i=1}^{N} w_k'(i)\, \hat{a}_{k/H_0}(i)\, y_k'(i)\right]^2 \underset{H_0}{\overset{H_1}{\gtrless}} T}{\frac{1}{2N}\sum_{k=1}^{2}\sum_{i=1}^{N} \left(w_k'(i)\right)^2 \sum_{k=1}^{2}\sum_{i=1}^{N} \left(\hat{a}_{k/H_0}(i)\, y_k'(i)\right)^2} \tag{14}$$

Interestingly, the Rao detector assumes the form of a normalized correlation detector, where the watermark shape is correlated with the normalized observations. One may recall that earlier the GLRT detector in its simplified form in Eq. (11) also assumes the form of a correlator. The Rao detector in Eq. (14) is invariant with respect to the strength of the watermark. The parameter estimates $\hat{a}_{k/H_0}(i)$ are also obtained by the MAP methodology as for the GLRT detector in Eq. (12).

5. NUMERICAL EXPERIMENTS

Numerical experiments are used to test the performance of the detectors based on the proposed image model. In order to establish statistical significance of our results we used 200 representative images (10 from each one of the 20 categories) of the Microsoft Image Recognition data base [10]. To quantify the power of the watermark in our experiments, the so-called watermark to document ratio (WDR) is used, which is defined as

$$WDR = 20\log_{10}\left(\frac{\|\gamma \mathbf{w}\|}{\|\mathbf{f}\|}\right) dB \,. \tag{15}$$

To quantify the detection performance, the receiver operating characteristics (ROC) curves are used. In particular, the area under the ROC (AUROC) curve for false alarm probability range [0-0.1] is used to quantify the performance of the detector at low false alarm rates; the total area under the ROC curve is also computed to quantify the overall performance of the detector. These two metrics are referred to as AUROC1 and AUROC2, respectively, in the rest of the paper.

ROCs curves were obtained using the same watermark which was added to all the 200 images of our data base. Then, the test statistic was evaluated for the 200 images with the watermark and the 200 images without the watermark.

For comparison purposes we considered detectors that are based on wavelet transform and GGD modeling. To the best of our knowledge, a non-adaptive wavelet model is typically used in the existing work in the literature, e.g., [2] and [4], where a single GGD model is assumed for all wavelet bands. In our experiments, we considered an "adaptive" wavelet GGD model, in which we used a

different GGD model for each wavelet band. For fairness to the wavelet GGD detectors, watermarking was performed in the wavelet domain of the images. The watermark was imbedded in the 2nd level of the discrete wavelet transform (DWT). The Daubechies-8 2-D separable filters were used. In all experiments the watermarked images were first quantized using 8 bits per pixel accuracy in the spatial domain before watermark detection.

In the experiments we tested the proposed detectors for a number of WDRs averaging over the set of 200 images as explained above. More, specifically, in Table 1 we summarize AUROC1 and AUROC2 results for different WDRs for the GLRT detectors using both GGD wavelet, and in Table 2 the Rao test using both GGD wavelet, and the proposed prior. In Figures 1 and 2 we show typical examples of ROCs obtained in these experiments. From the AUROC in Tables 1 and 2 and the ROCs in Figures 1 and 2 below the superiority of the detectors based on the proposed prior for this experiment is clear. It is worth noting here, for the sake of simplicity, the *same* value of the Gamma hyper prior parameter -*l*- was used for all test images. Thus, in this experiment the detectors based on the proposed prior have a handicap as compared to detectors where they are adapted to each image. In spite of this the proposed detectors proved superior to previous state-of-the-art detectors.

REFERENCES

[1] I. Cox, M. Miller, and J. Bloom, *Digital Watermarking*, Morgan Kaufman, 2002.

[2] Q. Cheng and T. S. Huang, "An additive approach to transform-domain information hiding and optimum detection structure," *IEEE Trans. on Multimedia*, vol. 3, no. 3, Sept. 2001.

[3] Q. Cheng and T. S. Huang, "Robust optimum detection of transform domain multiplicative watermarks*," IEEE Trans. on Signal Processing*, vol. 51, no. 4, April 2003.

[4] A. Nikolaidis and I. Pitas, "Asymptotically optimal detection for additive watermarking in the DCT and DWT domains," *IEEE Trans. on Image Processing*, vol. 12, no. 5, pp. 563-571, May 2003.

[5] G. Chantas, N. P. Galatsanos, and A. Likas, "Bayesian restoration using a new nonstationary edge preserving image prior", *IEEE Trans. on Image Processing*, Vol. 15, No. 10, pp. 2987-2997, October 2006.

[6] F-C. Jeng, J. W. Woods, "Compound Gauss-Markov fields for image estimation," *IEEE Trans. on Signal Proc*. vol. 39, no. 3, March 1991.

[7] J. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer Verlag, 1985.

[8] S. M. Kay, *Fundamental of Statistical Signal Processing: Detection Theory*, vol. 2 Prentice Hall, 1998.

[9] S. M. Kay, "Asymptotically optimal detection in incompletely characterized non-Gaussian noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, no. 5, pp. 627–633, May 1989.

[10] Microsoft Research Cambridge Object Recognition Image Database, http://research.microsoft.com/ downloads

Table 1. AUROCs for GLRT detectors

WDR dB	*(AUROC1, AUROC2) GGD Wavelet*	*(AUROC1, AUROC2) GLRT (proposed prior)*
-61	(0.0540, 0.9925)	(0.0862, 0.9680)
-62	(0.0166, 0.7048)	(0.0785, 0.9540)
-63	(0.0041, 0.6039	(0.0831, 0.9512)
-64	(0.0052, 0.5307)	(0.0763, 0.9238)
-65	(0.0014, 0.5051)	(0.0746, 0.9149)

Table2. AUROCs for RAO detectors

WDR	*(AUROC1,AUROC2) RAO GGD Wavelet*	*(AUROC1, AUROC2) RAO (proposed prior)*
-60	(0.0894, 0.9844)	(0.0993, 0.9992)
-61	(0.0675, 0.9679)	(0.0883, 0.9983)
-62	(0.0672, 0.9165)	(0.0775, 0.9848)
-63	(0.0348, 0.8001)	(0.0694, 0.9432)
-64	(0.0300, 0.5419)	(0.0389, 0.8106)

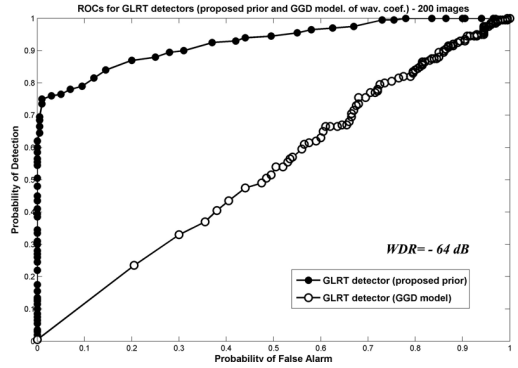

Figure 1. ROCs for GLRT detectors (proposed prior and GGD model), WDR=-64 dB

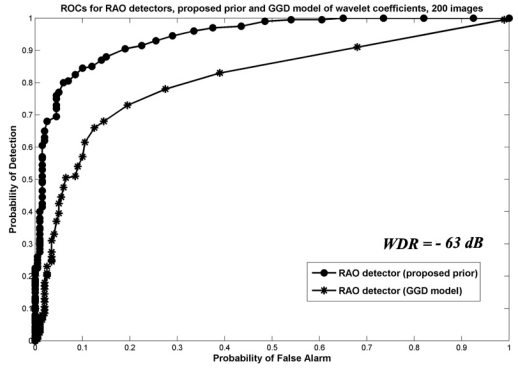

Figure 2. ROCs for RAO detectors (proposed prior and GGD model), WDR=-63 dB